THE RETRIEVAL PROCESS IN THE SAFRS SYSTEM WITH THE CASE-BASED REASONING APPROACH

Souad Demigha
Centre de Recherche en Informatique (C.R.I), Université de Paris 1 – Panthéon-Sorbonne
90 rue de Tolbiac 75634 Paris Cedex 13, France

Keywords: Retrieval process, SAFRS, case-based reasoning, training, MAP, intention, strategy.

Abstract: The paper presents the retrieval process in the SAFRS system (system supporting the training of radiologists-senologists) with the case-based reasoning approach (CBR, which is adopted to represent the experience of expert radiologists-senologists under the form of cases) and modelized with the MAP concept. The retrieval process relies on a procedure of case-based reasoning for retrieval of similar cases formalized using a MAP, a re-use methodology named the retrieval MAP. The model of the MAP is an intentional representation system. It is based on concepts of intention and strategy. The concept of intention (or a goal) aims to capture the objective to be achieved. A strategy is the manner an intention is achieved. The retrieval process with the MAP is a multi-step/multi-algorithm process, which permits to retrieve similar cases in various modes and strategies. It is achieved according to three complex strategies: global strategy (or global retrieval strategy), elementary strategy (or elementary retrieval strategy) and mixed strategy (or mixed retrieval strategy).

1 INTRODUCTION

The SAFRS system (Système d'Aide à la Formation des Radiologues-Sénologues) is a training system in the domain of radiology-senology. It aims at capitalizing and reusing the experience of radiologists-senologists in order to enable junior radiologists-senologists to have access to and learn from the experience of experts. Experts' experience is represented as knowledge; both *product knowledge (mammographies and associated diagnoses…)* and *process knowledge (heuristics)* are considered. While the product is the result to be achieved, the process is the way the result is achieved. The paper presents the retrieval process in the SAFRS system with the case-based reasoning approach (CBR) and which is modelized with the MAP concept representing the process knowledge. The case-based reasoning is adopted to represent the experience of expert radiologists-senologists as cases (Aamodt and al, 1994). This allows to obtain an oriented-object model with the UML formalism (Unified Modelling Language) structured as cases. The retrieval process is divided into three hierarchic levels: the first level (the case) is a patient at different intervals of treatment (time). A case may comprise several successive senologic episodes. The second level (the sub-case) is one senologic episode (clinical examination, image reading, radiological interpretation, and anatomo-pathological examination) for a given patient. The third level (the sub-sub-case) represents one phase of a senologic episode for a given patient (clinical examination OR image reading…), (Demigha and Prat, 2004).

The retrieval process using the MAP is a multi-step/multi-algorithm process, which permits to retrieve similar cases in various modes and strategies. It is achieved according to three complex strategies: global strategy (or global retrieval strategy), elementary strategy (or elementary retrieval strategy) and mixed strategy (or mixed retrieval strategy). The global strategy allows for retrieval at the global level, i.e. the case. The retrieval process starts at the sub-sub-case: it represents one phase of a senologic episode for a given patient (clinical examination OR image reading OR radiological interpretation OR anatomo-pathological examination), then we go to the intermediate level (the sub-case: it is one senologic episode for a given patient), until cases of interest in

Demigha S. (2007).
THE RETRIEVAL PROCESS IN THE SAFRS SYSTEM WITH THE CASE-BASED REASONING APPROACH.
In *Proceedings of the Ninth International Conference on Enterprise Information Systems - AIDSS*, pages 468-473
DOI: 10.5220/0002354904680473

the treatment of the new case (target case) are found; finally we aggregate at the case level. Elementary strategy allows combining one to three phases of the radiological process. The mixed strategy aims to go back and start from the elementary level (the sub-case) until finding cases of interest in the treatment of the new case (target case).

The paper structure is as follows:
- Section 2 provides in details the retrieval process with the concept of the MAP.
- Section 3 provides the evaluation of the retrieval process with the concept of the MAP.
- Section 4 is the conclusion with further research works in progress.

2 THE RETRIEVAL PROCESS IN THE SAFRS SYSTEM

In the SAFRS system, the retrieval process is modelled by a MAP called the retrieval MAP (see Figure 1), (Demigha, 2005).

2.1 The Model of the MAP

The model of the MAP is an intentional representation system. It is based on concepts of *intention* and *strategy*. It includes one or several *sections*. A *section* is based on two concepts: *intention (or goal)* and *strategy*. The concept of *intention* aims to capture the objective to be achieved at one time of a process. A *strategy* is the manner to achieve an intention. A *section* is an aggregation of two types of intentions: a *source intention*, a *target intention* and a *strategy* as well.

A MAP is represented by a graph oriented and labelled. Intentions represent nodes and strategies represent the arcs. A section is then represented by two nodes linked by an arrow. A section must be selected when it is initialized. The selection of sections is based on directives. A directive includes a *signature* and a *body*. A *signature* represents the visible part of the directive. A *signature* is defined by the couple *<(intention), situation>*. A *body* defines the followed step in order to satisfy the intention captured in the signature. A directive includes two types of directives: *strategic directive* and *tactical ones*: (1) the *strategic directive* represents a strategic view of the multi-step development based on a set of intentions and strategies. It is represented by a MAP and a set of associated directives and (2) the *tactical directive* has a three-structure. It is composed of three other directives (*context*: a *context* represents the development of a process by a hierarchy of contexts): *plan, selection* (the selection of several alternative sub-directives) and *executable*. *A plan directive* corresponds to a complex problem decomposed into a set of sub-problems. The execution of the composed directives is defined by a graph. The nodes of the graph are directives (components of the plan). Arcs (previous links) represent arranged or parallel transitions between directives. *A selection directive* corresponds to a situation that necessitates the exploration of different possibilities. *An executable directive* corresponds to an intention which can be characterized by an action of the product transformation or an action of selection of an other directive.

2.2 Similarities

In the senology case representation model, cases are collections of *objects*, each of which it is described by a set of *attribute-value pairs*. The structure of an object is described by an object class that defines the set of attributes together with a type (set of possible values or sub-objects) for each attribute. Object classes are arranged in a class hierarchy, that is, a tree in which sub-classes inherit attributes as well as their definition from the current class.

We define a hierarchy of attributes types. New types are defined by building sub-types of the existing elementary types shown in Table 1. They differ in their usability: A type may be used in an *immediate* or *derived type*. While *immediate types* cover the whole range of possible values of a type, *derived types* are restricted in their range by defining an enumeration of elements of their elementary types or, in case of numeric types, by specifying an interval (Bergmann and Althoff, 1998).

Table 1: Elementary types in the SAFRS system.

Type	**Usability**
Integer	Immediate and derived
Float	Immediate and derived
Date	Immediate and derived
Boolean	Immediate only
String	Immediate and derived
Enumeration	Immediate and derived
Ordered Enumeration	Derived only
Text	Derived only

The approach we have chosen to determine similarities is to establish a comparison between attributes (attribute by attribute), then to each attribute corresponds a comparison measure, it is a *local similarity measure*, it determines a similarity between two attribute values, and for each object we determine *a global similarity measure* which

determines the similarity between two objects (or between the case and the query) based on the local similarity of the belonging attributes. The local similarity measure allows to compare any two types values. It returns a numeric value from the interval [0..1]. This value is further used in the computation of a global similarity.

2.3 The Retrieval Process

The retrieval process is, with the MAP, a multi-step/multi-algorithm process, which permits to retrieve similar cases in various modes. The retrieval MAP of the SAFRS system represented on the graph (Figure 1) defines, besides the two intentions *'to start'* and *'to stop'*, two major intentions for the retrieval process *achievement 'to elaborate the new case'* and *'to retrieve similar cases'*.

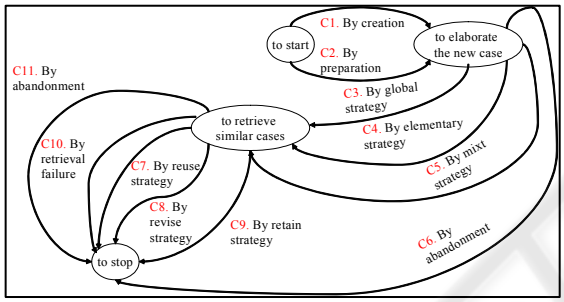

Figure 1: The retrieval MAP.

The intention *'to elaborate the new case'* is achieved according to two strategies: *'by preparation'* and *'by creation'*, contained in sections C1 and C2, respectively. It consists in describing a new case (new case to be diagnosed). From the new problem, the case is elaborated from the general knowledge, while keeping only relevant information. If the case description is complete, then the intention *'to elaborate the new case'* has the same meaning as the creation (strategy *'by creation'*) or the preparation (strategy *'by preparation'*) of the case, else the creation phase is simplified.

The intention *'to retrieve similar cases'* is achieved according to three complex strategies: *'global strategy' (or global retrieval strategy), 'elementary strategy'(or elementary retrieval strategy)* and *'mixed strategy' (or mixed retrieval strategy)* that are contained in sections C3, C4 and C5, respectively. The *'global strategy'* included in section C3 allows for retrieval *at the global level*, i.e. *the case*. The retrieval process starts at *the sub-sub-case level*, then we go to *the intermediate level, the sub-case*, and finally it ends to aggregate at the case level. '*Elementary* strategy' included in section C4 allows to combine one to three phases, i.e. *the sub-sub-case* ('image-reading', 'radiological interpretation' and 'anatomo-pathological examination') of the senological process. The *'mixed strategy'* included in section C5 allows to combine the first two strategies (global and elementary ones). It aims to go back and start from *the elementary level (sub-case),* until it finds cases of interest in the treatment of the new case (target case). The *'abandonment strategy'* included in section C6 allows the case expert to abandon his/her retrieval process for the new case, before starting the retrieval when he/she makes mistakes in his/her reasoning, thus allowing him/her to start again the retrieval process, without starting from the very beginning, i.e. from the source intention *'to start'* of the MAP.

Once the case expert has carried out the retrieval process, i.e. he/she succeeded or failed in searching an interesting case for solving the new problem, he/she has got four possibilities to treat this new case: the *'reuse strategy'* included in section C7 allows to revise the validity of retrieved solution, which is retained for the goal problem (new problem to solve). The *'revise strategy'* included in section C8 allows to revise the case according to three steps: to revise it *'by test'*, *'by correction'* and finally *'by validation'*. The *'retained strategy'* included in section C9 allows to integrate to the case base the new solved problem, if the latter confers novel abilities to the system. The strategy *'by retrieval failure'* included in section C10 allows to send back a negative result from the case base to the case expert when no case could be identified as similar enough to the target case (new case). Finally, the last strategy *'by abandonment strategy'* included in section C11 allows the case expert to abandon the retrieval of similar cases if he/she deems it necessary, even after the overall process is achieved.

We detailed in the three following sub-sections the three main strategies for the retrieval process.

2.3.1 The Global Strategy

The global retrieval strategy consists in retrieving the case in its totality. The Figure 2 models this strategy. Indeed, it is a *plan directive*: *<(new case), to research the similar cases by global strategy>* composed of a hierarchy of plans which contains three contexts: *plan, selection and executable.* The plan directive DRI_3 proposes three sub directives:

- $DRI_{3.1}$: *<(new case), to calculate similarities at the sub-sub-case level>**; (* means an iterative form); -$DRI_{3.2}$: *<(sub-sub-cases selected), to*

calculate similarities at the sub-case level>* ; - $DRI_{3.3}$: <*(sub-cases selected), to calculate similarities at the case level*>*.

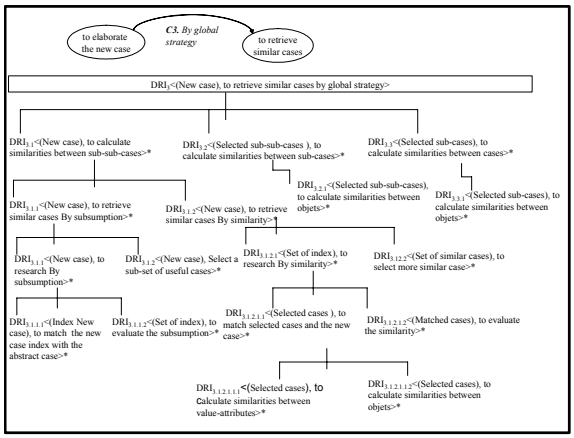

Figure 2: The global strategy (a plan directive: hierarchy of plan contexts).

As shown on Figure 2, the plan directive $DRI_{3.1}$: <*(new case), to calculate similarities between sub-sub cases*>* proposes two plan contexts for the realization of its intention: - $DRI_{3.1.1}$: <*(new case), to retrieve similar cases by subsumption*>* ; - $DRI_{3.1.2}$: <*(new case), to retrieve similar cases by similarity*>*.

The subsumption is a mechanism of discrimination. The directive $DRI_{3.1.1}$ is performed by the execution of two plan contexts:

- $DRI_{3.1.1}$: <*(new case), to retrieve similar cases by subsumption*>*. The intention 'to research by subsumption' is performed via two executable contexts: $DRI_{3.1.1.1}$: <*(index new case), to match the new case index with the abstract case*>* and $DRI_{3.1.1.2}$: <*(set of indices), to evaluate the subsumption*>*. To evaluate the subsumption consists of browsing a net of indices where, at each node, cases are selected by taking into account the subsumption criterion.

For facilitating the retrieval process, the case is abstract in order to extract indices. The abstraction is aimed to divide the problem descriptors of the input into two classes: the relevant descriptors (useful) and the non-relevant descriptors (not useful) or noises. The abstraction consists in eliminating noises.

- $DRI_{3.1.2}$: <*(new case), to select a sub-set of relevant cases*>*.

The intention *'to select a sub-set of relevant cases'* eliminates the very distant cases and selects a set of cases that are suitable for the target problem. It implies that cases are organized in a classification hierarchy according to relevant characteristics. The selection of these characteristics determines the capability to retrieve the 'best' cases.

After restricting the research space, the case author performs a more specific comparison between the target problem and each source case previously selected by discrimination *'by subsumption'* with the plan directive *'by similarity'*: $DRI_{3.1.2}$: <*(New case), to retrieve similar cases by similarity*>* .

The directive $DRI_{3.1.2}$: <(new case), to retrieve similar cases by similarity>* is performed by two plan contexts: $DRI_{3.1.2.1}$: <*(set of index), to research by similarity*>* and $DRI_{3.12.2}$: <*(set of similar cases), to select the most similar case*>*.

- *'To research by similarity'* (to research similar cases) performs a comparison more specific between the target problem and the source case previously selected by discrimination. This comparison necessitates a two by two comparison of cases, attribute by attribute. This directive proposes two plan directives for the realization of its intention $DRI_{3.1.2.1.1}$: <*(selected cases), to match selected cases and the new case*>* and $DRI_{3.1.2.1.2}$: <*(matched cases), to evaluate the similarity*>*.

- The intention *'to match selected cases and the new case'*: the matching process compares two by two characteristics of cases. In most systems, the matching is performed on characteristics of cases: it is a global matching (global similarity by attribute weighting at a local similarity level).

- The intention *'to evaluate the similarity'*: a similarity measure is used in order to arrange source cases by decreasing the similarity with the target case. The evaluation is performed by considering common characteristics; each one has a significant importance level (weight) of the role that each element of a problem plays in the reuse of elements of the solution. The similarity evaluation is assumed to depict the facility of the reuse of a source case.

- *'To select the most similar case'*: the solution of cases having the best 'score' is selected for the target problem. The directive plan $DRI_{3.1.2.1.1}$: <*(selected cases), to match selected cases and the new case*>* proposes two selection alternatives to complete the retrieval process: $DRI_{3.1.2.1.1.1}$ <*(selected cases), to calculate similarities between attributes*>* and $DRI_{3.1.2.1.1.2}$ <*(selected cases), to calculate similarities between objects*>*. These directives allow the computation of similarity measures between attribute-values *(a local similarity measure)* and objects *(global similarity measure) (sub-sub-case, sub-case and case)*.

2.3.2 The Elementary Strategy

The objective of *elementary strategy (or elementary retrieval strategy)* is to offer the case author various possibilities to resolve his/her problem. In the absence of complete information on the new case, the case author only considers into the case base the knowledge that resembles new knowledge. The case author can start the process with some knowledge of one phase of different phases of a patient x; for instance, this knowledge is compared with the knowledge of the new case. The case author selects knowledge of another phase concerning another patient y. He/she combines these knowledge and reiterates the process whenever required to make a diagnosis. All these fragments, coming from different phases of different patients or even from a same patient, combined together (in the case that the patient had previous reports), make up one solution of the new problem to solve. The assessment of the similarity (attributes and objects) is performed in the same manner as the global strategy and the mixed strategy.

The intention *'to calculate similarities between attribute-values'* of the directive $DRI_{3.1.2.1.1.1}$ allows to use the hierarchy of UML types (Demigha and Prat, 2004). Indeed, according to various types of attributes, a similarity measure is selected.

The two other main sub-directives of the DRI_3: $DRI_{3.2}$: *<(selected sub-sub-cases), to calculate similarities between sub-cases* and $DRI_{3.3}$: *<(selected sub-cases), to calculate similarities between cases>** are executable contexts and thus are not factorized.

- The second sub-directive of the directive $DRI_{3.2}$: *<(selected sub-sub-cases), to calculate similarities between sub-cases>** is a plan directive including one context plan: $DRI_{3.2.1}$: *<(selected sub-sub-cases), to calculate similarities between objects>**.

- The third sub-directive of the directive $DRI_{3.3}$: *<(selected sub-cases), to calculate similarities between cases >** is a plan directive including one context plan: $DRI_{3.3.1}$: *<(selected sub-cases), to calculate similarities between objects>**.

Figure 3 models this strategy. Indeed, it is a plan directive: *<(new case), to research similar cases by elementary strategy>* composed of a hierarchy of plans which include three contexts: plan, selection and executable.

The plan directive DRI_4 proposes three principal sub-directives: - $DRI_{4.1}$: *<(new sub-sub-case image reading phase), to calculate similarities at the image reading phase>** ;- $DRI_{4..2}$: *<(solution part of the image reading phase), to calculate similarities at the radiological interpretation level (RI)>** ; - $DRI_{4..3}$: *<(solution part of the RI phase), to calculate*

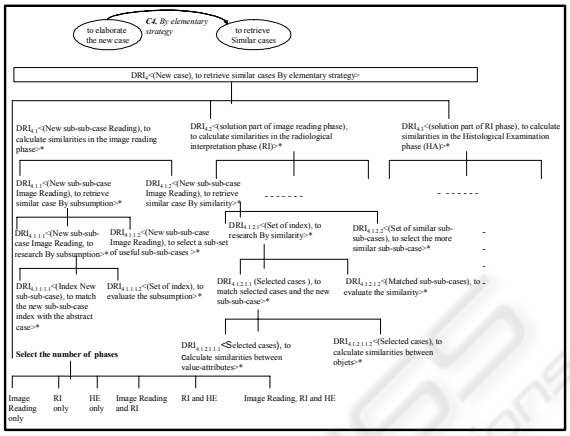

Figure 3: The elementary strategy (a plan directive: hierarchy of plan contexts).

*similarities at the anatomo-pathological examination phase (AE)>**.

2.3.3 The Mixed Strategy

The mixed strategy allows to combine the first two strategies, the global strategy and the elementary strategy. For a radiologist (in the case that he/she combines several knowledge from various sources), the interest of this strategy lies in picking up knowledge at the intermediate level, to find again archives of previous examinations, and thus to obtain a full knowledge.

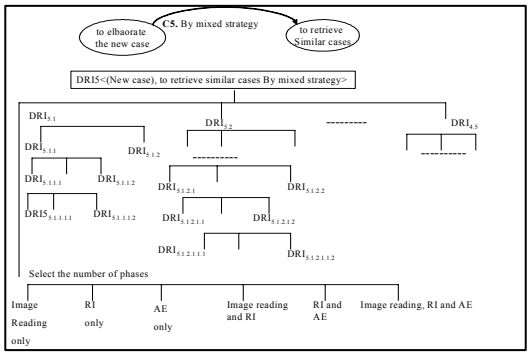

Figure 4: The mixed strategy (a plan directive: hierarchy of plan contexts).

Figure 4 models this strategy. Indeed, it is a plan directive: *<(new case), to research similar cases by mixed strategy>* composed with a hierarchy of plans including three contexts: plan, selection and executable.

The directive plan DRI_5 proposes five main sub-directives: - $DRI_{5.1}$: *<(new sub-sub-case image*

reading phase), to calculate similarities at the image reading phase>*; - $DRI_{5.2}$: <(solution part of the image reading phase), to calculate similarities at the radiological interpretation phase (RI)>*; - $DRI_{5.3}$: *<(part solution of the RI phase), to calculate similarities at the anatomo-pathological examination phase (AE);* - $DRI_{5.4}$: *<(selected sub-sub-cases), to calculate similarities at the sub-case level*>*; - $DRI_{5.5}$: *<(selected sub-cases), to calculate similarities at the case level)*>.

As shown on Figure 4, the directive $DRI_{5.1}$ is a hierarchy of directives of plan contexts, selection and executable. This hierarchy has the same course as the directive $DRI_{3.1}$ of the global strategy. We do not provide details of the steps of calculation of similarities. Extensive details for the retrieval process are presented in (Demigha, 2005).

3 DISCUSSION

This paper deals with our retrieval process of the case-based reasoning training system which we built based on the MAP model. The latter has several advantages: 1) as a process meta-model in the radiologists-senologists modelling approach to their interpretation, it enables, thanks to the directives, a fast and simple access to knowledge. Actually, the MAP offers a hierarchical and structuring approach using selection and mixed strategies, 2) thanks to these strategies, the radiologists-senologists can at the same time have a free and diversified access in order to browse dynamically the MAP. Selecting a strategy is made as the realization of the intentions is carried out. This means that the selection is a dynamic process and the construction of the paths is achieved according to the situations that are met with, 3) most of the time; the radiologist-senologist does not have enough knowledge in data-processing and solves randomly his/her daily problems. He/she has the advantage of adapting to the intentional reasoning of the MAP. At no point is the radiologist-senologist (especially the junior radiologist) forced to carry out a particular intention or to apply a strategy of realization of particular intention, unless otherwise required by the situation, 4) the intentional approach is structuring. Thanks to the intentions, it makes it possible to synthesize and to abstract the details in order to concentrate on the most important at even the priorities, 5) there are various ways of carrying out the intentions thanks to the strategies; as a result, the problems can be solved in a flexible and versatile way and 6) the strategies enabled us to provide the radiologists-senologists with the heuristics enabling them to use the best cases. These heuristics are strongly related to the complex structure and to the dependence on the cases, therefore allowing for a lowering of the cost of adaptation.

4 CONCLUSIONS

In this paper, we have presented the retrieval process in the SAFRS system with the case-based reasoning approach using the MAP model. The retrieval process in the SAFRS system we developed has an original aspect: this approach relies on the formal description of the process in an intentional manner.

It is a complex process. It describes in a quite accurate and detailed fashion the way we retrieve similar cases according to various modes. The exploitation of the layered case structure allows for the search of a similar case by composition of sub and sub-sub similar cases and by a set of a powerful similarity measures. Implementation and validation of the retrieval process will be developed in an other paper.

REFERENCES

Aamodt, A., Plaza, E., Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, 1994. In *AI Communications*, 94' volume 7, number 1.

Bergmann, R, Althoff. K.D, Methodology for Building CBR Applications, 1998. In *M.Lenz, B.Bartsh-Sport, H.-D. Burkhard, & S. Wess (Eds). Case-Based Reasoning from Foundations to Applications*, Spring-Verlag, 98'.

Demigha, S., Prat, N., A case-based training system in radiology-senology. 2004. In *Information and Communication Technologies: from Theory to Applications (ICTTA) 04'*, Damascus, Syria.

Demigha, S., Un système à base de connaissances en radiologie-sénologie dédié à la formation des radiologues-sénologues juniors, 2005. *Phd thesis in Informatic*, Université de Paris 1 Panthéon-Sorbonne.